Some Remarks On Exchangeable Normal Variables And Potential Applications

by

Seymour Geisser[1]

University of Minnesota

Technical Report No. 441

July 1984

[1]This work was supported in part by NIGMS 25271

1. Introduction.

Consider the following apparently simple problem:

Suppose we have a set of independent variables $Z_1,\ldots,Z_{N+1},\ldots,Z_{N+n}$ which are $N(\mu,\sigma^2)$. The variables $Z_1,\ldots,Z_N$ will be observed and particular functions of the unobserved set $Z_{N+1},\ldots,Z_{N+n}$ are to be predicted. Let $\bar{Z}=N^{-1}\sum_{i=1}^{N} Z_i$ and $X_i=(Z_{N+i}-\bar{Z})/\sigma\sqrt{1+N^{-1}}$. Then $X_1,\ldots,X_n$ are multivariate normal with marginal distribution for X_i being $N(0,1)$ and with common correlation coeficient $\frac{1}{N+1}$. Suppose we want to predict R, the number of Z_{N+i}'s greater than $\bar{Z}$. We present some results that would to a degree characterize the distribution of R when $X_1,\ldots,X_n$ are exchangeable and $\rho\in[0,1]$, and then proceed to more general problems of this type.

Theorem Let R be the number of a set of exchangeable normal variables $X_1,\ldots,X_n$, that exceed their mean. If ρ is the common correlation coefficient among the X_i's then the probability function of R is: (1) symmetric, (2) strictly increasing until $\left[\frac{n}{2}\right]$ and strictly decreasing starting from $\left[\frac{n}{2}\right]$ or $\left[\frac{n}{2}\right]+1$ (depending on whether n is even or odd) for $0\leq\rho<\frac{1}{2}$, (3) uniform for $\rho=\frac{1}{2}$, (4) strictly decreasing until $\left[\frac{n}{2}\right]$ and then strictly increasing starting from $\left[\frac{n}{2}\right]$ or $\left[\frac{n}{2}+1\right]$ for $\frac{1}{2}<\rho<1$, (5) as $\rho\to 1$, all the probability is concentrated equally at $R=0$ and $R=n$, (6) for any fixed non-negative integer r_0 no larger than $\left[\frac{n}{2}\right]$, $\Pr[R\leq r_0|\rho]$ increases with ρ, (7) for $R=0$ or n, the probability function is an increasing function of ρ, for $R=r\neq 0$ or n it is a decreasing function of ρ if $(n+1)\binom{n}{r}>2^n$ and if the previous inequality is reversed it increases monotonically until $\rho = 1/2$ and then decreases as $\rho \to 1$.

Proof:

It is no loss of generality to assume, $E(X_i) = 0$, $\mathrm{Var}(X_i) = 1$, and $\mathrm{Cov}(X_i, X_j) = \rho \geq 0$. Now due to the exchangeability assumption

$$\Pr[R = r|\rho] = \binom{n}{r} \Pr[X_1 > 0, \ldots, X_r > 0; X_{r+1} < 0, \ldots, X_n < 0].$$

Further it is clear that we can decompose X_i as

$$X_i = Y_i - V \qquad i = 1,2,\ldots n$$

where Y_i's are i.i.d. $N(0, 1-\rho)$ and independent of V which is $N(0,\rho)$.

Hence

$$\Pr[R = r|\rho] = E_V\left[\binom{n}{r} \Pr\{Y_i > V, \ldots, Y_r > V; Y_{r+1} < V, \ldots, Y_n < V|V\right]$$

$$= \binom{n}{r} \int_{-\infty}^{\infty} \Phi^{n-r}\left(\frac{v}{\sqrt{1-\rho}}\right) [1 - \Phi\left(\frac{v}{\sqrt{1-\rho}}\right)]^r \frac{e^{-\frac{v^2}{2\rho}}}{\sqrt{2\pi\rho}} dv \tag{1.1}$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Let $v = y\sqrt{\rho}$ and $\sqrt{\rho/(1-\rho)} = \tau$ then

$$\Pr[R=r|\tau] = \binom{n}{r} \int_{-\infty}^{\infty} \Phi^{n-r}(y\tau) \; [1-\Phi(y\tau)]^r d\Phi(y) \tag{1.2}$$

To establish symmetry i.e. $\Pr[R=r|\tau] = \Pr[R=n-r|\tau]$ we write the integral as a sum of two integrals,

$$\Pr[R=r|\tau] = \binom{n}{r}[I_1 + I_2]$$

where

$$I_1 = \int_{-\infty}^{0} \Phi^{n-r}(y\tau)[1 - \Phi(y\tau)]^r \, d\Phi(y)$$

$$I_2 = \int_0^{\infty} \Phi^{n-r}(y\tau)[1 - \Phi(y\tau)]^r \, d\Phi(y).$$

By letting $y=-x$ and recalling that $\Phi(-a)=1-\Phi(a)$, we obtain

$$I_1 = \int_0^{\infty} [1-\Phi(x\tau)]^{n-r}\Phi^{r}(x\tau)\,d\Phi(x) ,$$

$$I_2 = \int_{-\infty}^{0} [1-\Phi(x\tau)]^{n-r}\Phi^{r}(x\tau)\,d\Phi(x) .$$

Hence the integral is symmetric in r and n-r and

$$\Pr[R=n-r|\tau] = \Pr[R=r|\tau].$$

In particular for $\rho=0$, so that $\tau=0$, we obtain

$$\Pr[R=r[\rho=0] = \binom{n}{r}\frac{1}{2^n} ,$$

and for $\rho=\frac{1}{2}$, or $\tau=1$,

$$\Pr[R=r|\rho=\tfrac{1}{2}] = \binom{n}{r}\int_{-\infty}^{\infty} \Phi^{n-r}(y)[1-\Phi(y)]^{r}d\Phi(y) = \frac{1}{n+1} .$$

Further,

$$\lim_{\rho\to 1} \Pr[R=0|\rho] = \lim_{\rho\to 1} \Pr[R=n|\rho] = \frac{1}{2}$$

$$\lim_{\rho\to 1} \Pr[R=r|\rho] = 0 \text{ for } 1\leq r\leq n-1.$$

We now need to show that for $r\leq\left[\frac{n}{2}\right]-1$ and $\tau<1$, i.e., $\rho<\frac{1}{2}$

$$\Pr[R=r|\tau] < \Pr[R=r+1|\tau] \tag{1.3}$$

or

$$\binom{n}{r}\int_{-\infty}^{\infty} \Phi^{n-r}(y\tau)[1-\Phi(y\tau)]^{r}d\Phi(y) < \binom{n}{r+1}\int_{-\infty}^{\infty} \Phi^{n-r-1}(y\tau)[1-\Phi(y\tau)]^{r+1}d\Phi(y)$$

or equivalently

$$\int_{-\infty}^{\infty} \Phi^{n-r-1}(y\tau)[1-\Phi(y\tau)]^r[(n+1)\Phi(y\tau)-(n-r)]d\Phi(y) < 0. \qquad (1.4)$$

To verify the above directly requires a good bit of analysis but a heuristic demonstration can be given as follows: clearly

$$\Pr[R=r|Y\tau] < \Pr[R=r+1|Y\tau] \qquad (1.5)$$

$$\text{for} \quad r < n\Phi(Y\tau) - 1 \quad .$$

Hence taking the expectations with respect to Y on both sides of the two equations is indicative of the result (1.3) for

$$r < \frac{n}{2} - 1$$

$$\text{since} \quad E_Y[\Phi(Y\tau)] = \frac{1}{2} \quad \text{for all } \tau \ .$$

To demonstrate that $\Pr(R \leq r_0|\tau)$ is an increasing function of ρ or τ for $r_0 < \left[\frac{n}{2}\right]$ we note that the distribution function

$$\Pr\ [R \leq r_0|\tau] = \sum_{r=0}^{r_0} \binom{n}{r} \int_{-\infty}^{\infty} \Phi^{n-r}(y\tau)[1-\Phi(y\tau)]^r \, d\Phi(y)$$

$$= \sum_{r=0}^{r_0} \binom{n}{r} \int_{-\infty}^{\infty} \left\{\Phi^r(y\tau)]1-\Phi(y\tau)]^{n-r} + \Phi^{n-r}(y\tau)[1-\Phi(y\tau)]^r\right\} d\Phi(y)$$

$$= (n-r_0)\binom{n}{r_0}\int_0^{\infty}\left[\int_0^{1-\Phi(y\tau)} t^{n-r_0-1}(1-t)^{r_0}dt + \int_0^{\Phi(y\tau)} t^{n-r_0-1}(1-t)^{r_0}dt\right] d\Phi(y)$$

Taking the derivative of the above w.r.t. τ we obtain

$$\frac{dP}{d\tau} = (n-r_o) \binom{n}{r_o} \int_0^{\infty} y\phi(y\tau) \left[\Phi^{n-r_o-1}(y\tau)(1-\Phi(y\tau))^{r_o} - \Phi^{r_o}(y\tau)(1-\Phi(y\tau))^{n-r_o-1}\right] d\Phi(y)$$

The integrand is positive almost everywhere in the domain of integration since $\Phi(y\tau) > \frac{1}{2}$ for all $y\tau > 0$. Thus the derivative is positive and the result established. Result (7) can be demonstrated by taking the derivative of the probability function with respect to τ. The symmetry of the probability function also implies that, for every $n \geq 2$ and $1 \leq r \leq n - r$,

$$\Pr\,[r \leq R \leq n - r \mid \rho]$$

is a decreasing function of ρ.

2. Some Consequences and Other Characteristics of the Distribution

As noted above the theorem indicates that as ρ increases from 0 to $\frac{1}{2}$ the shape of the probability function changes from an inverted bowl to one which is uniform. As ρ goes from $\frac{1}{2}$ to 1, the probability function changes from uniform to bowl shaped. This implies that for maximizing a probability prediction with respect to a given number of integers, the symmetric center is appropriate for $\rho < \frac{1}{2}$.

Certain other characteristics of the distribution of R are easy to calculate e.g.

$$E(R|\rho) = \frac{n}{2}\ ; \qquad \mathrm{Var}(R|\rho) = \frac{n}{4} + \frac{n(n-1)\ \arcsin \rho}{2\pi} \tag{2.1}$$

We note that

$$E(R) = n\Pr[X_1 > 0] = n/2$$

and

$$\mathrm{Var}(R) = \frac{n}{4} + n(n-1)\ \{\Pr[X_i > 0, X_j > 0] - 1/4\}$$

where

$$\Pr(X_i > 0, X_j > 0) = \int_0^{\infty} \int_0^{\infty} f(x_i, x_j)\ dx_i dx_j = I$$

where (X_i, X_j) is $N(0,0,1,1,\rho)$. Now, by transforming $(X_i + X_j) = \sqrt{2}\ Z$ and $(X_i - X_j) = \sqrt{2}\ W$

$$I = \int_0^{\infty} f(z) \int_{-z}^{z} f(w)\ dw\, dz$$

where Z and W are independently and normally distributed with zero means and variance $1+\rho$ and $1-\rho$, respectively. Further transformation yields

$$I = \frac{1}{\pi} \int_0^{\infty} e^{-y^2/2} \int_0^{y\sqrt{\frac{1+\rho}{1-\rho}}} e^{-\frac{1}{2}x^2}\ dx\, dy$$

and then a transformation to polar coordinates i.e.

$$x = r \cos \theta$$

$$y = r \sin \theta$$

yields

$$I = 2 \int_0^{\infty} \frac{re^{-r^2/2}}{2\pi}\ dr \int_{\mathrm{Tan}^{-1}\sqrt{\frac{1-\rho}{1+\rho}}}^{\pi/2} d\theta = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}$$

and the result is obtained. We note also that for $n = 2$, we can obtain the exact probability function of R namely

$$\Pr[R = 0|\rho] = \Pr[R = 2|\rho] = \frac{1}{4} + \frac{\arcsin \rho}{2\pi}$$

$$\Pr[R = 1|\rho] = \frac{1}{2} - \frac{\arcsin \rho}{\pi} \quad .$$

For this case the m.l.e. of ρ is $\hat{\rho}=0$ if $R=1$ and $\hat{\rho}=1$ if $R=0$ or 2 and the probability function of $\hat{\rho}$ is obvious. Note that Var(R) given by (2.1) is an increasing function of ρ, $0 \le \rho \le 1$ with minimum variance for $\rho=0$ and maximum at $\rho=1$. For this distributional family the variance is misleading as an indicator of how well R can be predicted in the range $\frac{1}{2} \le \rho \le 1$. One would normally assume that prediction should improve as the variance decreases. Nothing could be further from the truth in this case. Obviously the best prediction occurs when $\rho=1$, where the variance is maximized, all points $1 \le R \le n-1$ are excluded with probability 1. Here one is certain to be correct with a 2 point predictor i.e. 0 and n. In general, once ρ exceeds $\frac{1}{2}$ the best probability predictors are equal tails of the distribution of R.

Since $E\left(\frac{R}{n}\right)=\frac{1}{2}$ and $\mathrm{Var}\left(\frac{R}{n}\right)=\frac{1}{4n}+\frac{(n-1)\arcsin\rho}{2\pi n}$, then

$$\lim_{n\to\infty} \mathrm{Var}\left(\frac{R}{n}\right)=\frac{\arcsin\rho}{2\pi}.$$

By equation (1.2) and the theorem of de Finetti (1937) we can calculate the limiting distribution of $\frac{R}{n}$. Set $\Theta = \lim_{n\to\infty}\frac{R}{n}$ then Θ has the distribution of $1-\Phi(Y\tau)$ where Y is N(0,1). Hence from

$$\Pr[\tau Y \le \tau y] = \Phi(y)$$

or

$$\mathrm{PR}[1-\Phi(\tau Y) \ge 1-\Phi(\tau y)] = \Phi(y)$$

we obtain

$$\Pr[\Theta \ge \Phi(-y\tau)] = \Phi(y) \qquad -\infty < y < \infty$$

or

$$\Pr[\Theta \leq \Phi(-y\tau)] = 1 - \Phi(y),$$

Hence the distribution for Θ can be easily calculated from normal tables. Note that for:

$$\tau = 0\,, \quad \Pr[\Theta = 1/2] = 1$$

$$\tau = 1\,, \quad \Pr[\Theta \leq \Phi(-y)] = \Phi(-y) \quad \text{i.e. uniform}$$

$$\tau \to \infty\,, \quad \Pr[\Theta = 0] = \Pr[\Theta = 1] = 1/2.$$

3. A More General Problem

R, the number of Z_{N+i}'s greater than $\bar{Z}$, has probability function $\Pr(R = r|\frac{1}{N+1})$ which is of type (2) in the theorem for $N > 1$, irrespective of σ^2. Note that for $N = 1$, $\Pr(R = r|\frac{1}{2}) = \frac{1}{n+1}$, which can be readily obtained directly. As $N \to \infty$ the limiting probability of R is

$$\binom{n}{r} \frac{1}{2^n}$$

which could be determined directly.

A more general and much more important case is where we calculate for the exchangeable set $X_1, \ldots, X_n, \ldots$

$$\Pr[X_i > x] = \Pr\left[\frac{Y_i}{\sqrt{1-\rho}} > \frac{V+x}{\sqrt{1-\rho}}\right]$$

where $Y_1, \ldots, Y_n$ and V are independent as in section 1, but $x \neq 0$. Hence if R is the number of X_i's that exceed x then

$$\Pr[R=r|\rho] = \binom{n}{r} \int_{-\infty}^{\infty} \Phi^{n-r}\left(\frac{y\sqrt{\rho}+x}{\sqrt{1-\rho}}\right)\left[1 - \Phi\left(\frac{y\sqrt{\rho}+x}{\sqrt{1-\rho}}\right)\right]^r d\Phi(y) \qquad (3.1)$$

Clearly for $x \neq 0$, $\Pr[R=r|\rho]$ is no longer symmetric. It is conjectured however that properties similar to that of (2) and (4) in the theorem of section 1 hold i.e. there is a ρ depending on x say, $\rho(x,n) = \rho^*$, such that the probability function is strictly increasing until $\left[n(1-\Phi(x)\right]$ for $0 < \rho < \rho^*$ and decreasing thereafter and vice-versa for $\rho^* < \rho < 1$ but not uniform at $\rho = \rho^*$. It is easy to show that as $\rho \to 1$ all of the probability tends to be concentrated at $R=0$ and $R=n$ such that $\lim_{\rho\to 1} \Pr(R=0) = \Phi(x)$ and $\lim_{\rho\to 1} \Pr(R=n) = 1-\Phi(x)$. A property similar to (6) is also conjectured for $r_0 \leq [n(1-\Phi(x)]$.

In the introductory example we were interested in the number of $Z_{N+1},\ldots,Z_{N+n}$ that were greater than $\bar{Z}$, and in that case knowledge of σ^2 was irrelevant. However, if we are interested in the number greater than some given z, σ^2 is relevant since $X_i \sim N(0,1)$

$$\Pr\,[Z_{N+i} > z] = \Pr\,[X_i > x] \tag{3.2}$$

where

$$x = \frac{z-\bar{z}}{\sigma\sqrt{1+N^{-1}}} \;.$$

When σ^2 is unknown, set

$$X_i = \frac{Z_{N+i}-\bar{Z}}{s(1+N^{-1})^{1/2}} \tag{3.3}$$

where $s^2 = (N-1)\sum_{i=1}^{N} (Z_i - \bar{Z})^2$ and

$X_1,\ldots,X_n$ are jointly distributed as a multivariate student distribution centered at the origin with $E(X_i)=0$ if $N>2$ and $\mathrm{Var}(X_i)=\frac{N-1}{N-3}$, $\mathrm{Cov}(X_iX_j)=\frac{N-1}{(N-3)(N+1)}$ if $N>3$. A "best" multivariate normal approximation to this distribution is the one with these characteristics as measured by a Kullback-Leibler distance. Hence an approximate or large sample solution to the problem posed here involves the calculation indicated in (3.1) with x replaced by $x(N-3)^{1/2}(N-1)^{-1/2}$. Of course for any given $n > 1$, the calculation of (3.1) can be formidable.

Asymptotic results are available. Since for any given z by de Finetti's theorem

$$\lim_{n\to\infty} \frac{R}{n} = \Theta = \Pr[Z > z|\mu,\sigma^2] = 1 - \Phi\left(\frac{z-\mu}{\sigma}\right) \tag{3.4}$$

where Θ is a random variable whose distribution may be calculated. If we assume the usual vague prior for μ and σ^2

$$p(\mu,\sigma^2) \propto 1/\sigma^2 \tag{3.5}$$

then a posteriori the joint density of μ and σ^2 can be obtained from the fact that conditional on σ^2,

$$\mu \sim N\left(\bar{z},\ \frac{\sigma^2}{N}\,|\sigma^2\right)$$

and the marginal density of

$$\frac{(N-1)s^2}{\sigma^2} \sim \chi^2_{N-1} \ .$$

Using these results it is not difficult to show that if σ^2 is known,

$$\begin{aligned}
\Pr[\Theta \le \theta] &= \Pr\left[1 - \Phi\left(\frac{z-\mu}{\sigma}\right) \le \theta\right] \\
&= \Pr\left[1 - \theta \le \Phi\left(\frac{z-\mu}{\sigma}\right)\right] \\
&= \Pr\left[\mu \le z - \sigma\,\Phi^{-1}(1-\theta)\right] \\
&= \Phi\left\{\sqrt{N}\left(\frac{z-\bar{z}}{\sigma} - \Phi^{-1}(1-\theta)\right)\right\} \ .
\end{aligned} \tag{3.6}$$

When σ^2 is unknown then

$$\Pr[\Theta \leq \theta] = \Pr\left[1 - \Phi\left(\frac{z-\mu}{\sigma}\right) \leq \theta\right] = \Pr\left[\Phi\left(\frac{\mu - z}{\sigma}\right) \leq \theta\right]$$

$$= \Pr\left[\frac{\mu-z}{\sigma} \leq \Phi^{-1}(\theta)\right] = F_\beta\left(\Phi^{-1}(\theta)\right) \tag{3.7}$$

where $\beta = \frac{\mu-z}{\sigma}$ has distribution function $F_\beta(\cdot)$. After some analysis the density function of β can be obtained, Geisser (1967),

$$f(\beta) = \frac{\sqrt{N}\ e^{-N\beta^2/2}}{\sqrt{2\pi}\ \Gamma\left(\frac{N-1}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\sqrt{2}\beta N d\right)^j \Gamma\left(\frac{N+j-1}{2}\right)}{j!\ \left(1+Nd^2\right)^{\frac{N+j-1}{2}}} \tag{3.8}$$

where

$$d = \frac{(\bar{z} - z)}{(N-1)^{1/2} s} \quad . \tag{3.9}$$

It is noted that even for the asymptotic distribution of $\frac{R}{n} \to \theta$, the density of a very convenient transform of Θ, namely β, must be expressed as an infinite series - so that explicit results for finite n must be rather complicated. By calculating the characteristic function of β it is not difficult to show that the distribution of β tends to normality as N grows with mean and variance

$$\bar{\beta} = \frac{\sqrt{2}\ \Gamma(N/2)}{\Gamma\left(\frac{N-1}{2}\right)} d \doteq \frac{\bar{z}-z}{s} = k \tag{3.10}$$

$$\sigma_\beta^2 = N^{-1} + k^2 \left[1 - \frac{2\Gamma^2(N/2)}{(N-1)\Gamma^2\left(\frac{N-1}{2}\right)}\right] \doteq N^{-1} + \frac{k^2}{2(N-1)} \tag{3.11}$$

Hence

$$\Pr\left[\Theta \leq \theta\right] = \Pr\left[\beta \leq \Phi^{-1}(\theta)\right]$$

$$= \Pr\left[\frac{\beta-\bar{\beta}}{\sigma_\beta} \leq \frac{\Phi^{-1}(\theta) - \bar{\beta}}{\sigma_\beta}\right]$$

$$\doteq 1 - \Phi\left[\frac{\bar{\beta} - \Phi^{-1}(\theta)}{\sigma_\beta}\right] = 1 - \Phi\left[\frac{\bar{\beta}+\Phi^{-1}(1-\theta)}{\sigma_\beta}\right] \tag{3.12}$$

The latter should serve as a reasonably good approximation for $N \geq 30$.

We note that exact values for (3.7) can be obtained from the distribution of β or from the relation

$$\Pr[\Theta \leq \theta] = \Pr[T \leq -k\sqrt{n}] \tag{3.13}$$

where T is the "non-central student t" variate with n-1 degrees of freedom and non-centrality parameter $\sqrt{n}\Phi^{-1}(1-\theta)$, Aitchison (1964), Guttman (1970).

4. Examples

A user of ball bearings is planning to purchase 1000 whose advertised diameter is 8(mm). A random sample of 30 yields a mean diameter (mm), $\bar{z}=8.056$ with $s=.0928$. It is necessary that the diameter (mm) of a ball bearing exceed $z=7.9$ for its efficient use. The user will be satisfied with his purchase if he has a high degree of assurance that at least 90% of his purchase can be efficiently used. We now compute the probability, based on the previous normal assumptions with the specified vague prior on μ and σ, that at least 900 out of the 1000 will meet the need for efficient use.

Now from the non-central t distribution we calculate the exact asymptotic value of $\Pr(\Theta > .9|7.9) = .897$. Using the normal approximation to the exact asymptotic value we obtain

$$\Pr[\Theta > .9|7.9) \doteq \Phi\left[\frac{\Phi^{-1}(.1)+\bar{\beta}(7.9)}{\sigma_\beta(7.9)}\right] = .918$$

where from the data

$$\bar{\beta}(7.9) \doteq \frac{\bar{z} - 7.9}{s} = 1.6810 \qquad \sigma_\beta(7.9) \doteq .2825.$$

Given that $n = 1000$, it is very likely that the exact asymptotic value .897 is a highly accurate approximation to the exact value but if n were considerably less, 100 say, and 90 or more was required i.e. the same proportion, then it is likely that the exact value could drop by about .03. Indications of this appear in Geisser (1982) for exponentially distributed random variables where exact asymptotic results, with corrections for continuity, were used for the finite case. When compared with the exact finite results considerable descrepancies appeared for n even as large as 100.

This work can be extended to multiple components. We shall illustrate this with two components. Suppose that a piece of equipment, in order to function properly, requires both of two different components to exceed prescribed values. Suppose independent sets of data of size N_1 and N_2 are available on Component X_1 and Component X_2, and X_{ij} is independently distributed as $N(\mu_i, \sigma_i^2)$. Suppose a potential buyer of n of each type of component wants to calculate the probability that at least t out of n pairs will enable the equipment to function properly. Assume that n is sufficiently large so that the distribution of $\theta_i = \lim_{n\to\infty} \frac{R_i}{n}$ is a reasonable approximation to the distribution of R_i/n. Letting $Pr[Z_i > z_i | \mu_i, \sigma_i] = \theta_i$, and making the same assumptions as before we then generate two independent densities of the type given by (3.8)

$$f(\beta_1, \beta_2) = f(\beta_1 | N_1, d_1)\ f(\beta_2 | N_2, d_2) \qquad (4.1)$$

where

$$\beta_i = \frac{\mu_i - z_i}{\sigma_i}, \quad d_i = \frac{\bar{z}_i - z_i}{s_i \sqrt{N_i - 1}} \quad ,$$

and $s_i^2 = (N_i - 1)^{-1} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2$. Hence as in (3.10) - (3.12) for $i = 1, 2$

$$\bar{\beta}_i \doteq \frac{\bar{z}_i - z_i}{s_i} = k_i$$

$$\sigma^2_{\beta_i} \doteq N_i^{-1} + \frac{k_i^{\ 2}}{2(N_i-1)}$$

$$\Pr[\Theta_i > \theta_i] \doteq \Phi\left[\frac{\bar{\beta}_i - \Phi^{-1}(\theta_i)}{\sigma_{\beta_i}}\right] \tag{4.2}$$

Since

$$\Pr[\Theta_i > \theta_i;\ \Theta_2 > \theta_2] = \Pr[\Theta_1 > \theta_1]\ \Pr[\Theta_2 > \theta_2]$$

we can calculate for exceedance values z_1 and z_2,

$$\Pr[\Theta_1 > \theta_1;\ \Theta_2 > \theta_2 | z_1, z_2] = \Pr[\theta_1 > \theta_1 | z_1]\ \Pr[\Theta_2 > \theta_2 | z_2]$$

Assuming that the two components, one from X_1 and one from X_2, are chosen at random then the chance of obtaining T pairs of components, each member being within its stipulated tolerance, is

$$\Pr[T = t | R_1 = r_1, R_2 = r_2] = \frac{\binom{r_1}{t}\binom{n-r_1}{r_2-t}}{\binom{n}{r_2}} \tag{4.3}$$

for $t \leq \min(r_1, r_2)$. To obtain the probability that at least t out of the n pairs of components will enable the equipment to function properly we may approximate the distribution of $n^{-1}T = \Lambda$, conditional on θ_1 and Θ_2, as $N(\theta_1\theta_2,\ n^{-1}\theta_1\theta_2(1-\theta_1)(1-\theta_2))$. We then could calculate

$$\Pr[\Lambda > \lambda] = \int_\lambda^\infty \int_\lambda^\infty G(\lambda | \theta_1, \theta_2)\ dF(\theta_1 | z_1)\ dF(\theta_2 | z_2)$$

where $\quad G(\lambda | \theta_1, \theta_2) = 1 - F(\lambda | \theta_1, \theta_2)$

and F represents the appropriate distribution function. Note that this is still a difficult calculation, even though we have used asymptotic distributions for $\frac{R_1}{n}$, $\frac{R_2}{n}$, and $\frac{T}{n}$.

Another possibility is to approximate the unconditional mean and variance of T. Recalling that $\Theta_i = \Phi(\beta_i)$ we obtain

$$E(T) = E\left[\frac{r_1 r_2}{n}\right] \doteq n\, E[\Theta_1\Theta_2] \doteq n\, \Phi(\bar{\beta}_1)\,\Phi(\bar{\beta}_2)$$

$$\mathrm{Var}\ (T) \doteq n\Phi(\bar{\beta}_1)\left(1-\Phi(\bar{\beta}_1)\right)\Phi(\bar{\beta}_2)\left(1-\Phi(\bar{\beta}_2)\right) .$$

Using a normal approximation i.e.

$$\frac{T-E(T)}{\sqrt{V(T)}} \to N(0,1)$$

we can tentatively calculate the chance that $T > t_o$, but reasonable accuracy even for large values of n cannot be assured.

5. Remarks.

As n grows the exact limiting distribution of $Rn^{-1} \to \Theta$ can be found from its transform β. The normal approximation to the exact distribution of β which depends on N should be fairly accurate for $N \geq 20$. However the use of even the exact asymptotic distribution for R/n, will depend to a large degree on the magnitude of n. It is conjectured, on the basis of previous experience with the simple exponential sampling distribution Geisser (1982), that n will have to be fairly large, say $n \geq 100$, before the exact asymptotic distribution is an adequate approximation for the exact finite distribution of R/n. Further in the second illustration another distribution, the hypergeometric, is approximated by a normal which is probably reasonably adequate even for moderate n, say $n \geq 50$. However, in this illustration the fact that three successive approximate distributions are used may make one somewhat uneasy about the accuracy of the final result even if one has reasonable confidence in each of the successive approximations.

In order to make useful predictive probability calculations from a simple normally distributed process, a good deal of highly intricate calculations are necessary even when asymptotic results are used to approximate discrete and finite random variables. Clearly a combination of high speed computing and good numerical analytic techniques would be necessary to make all our calculations highly accurate.

6. Further Work

An important extension of this type of work is when Z lies in an interval that is not necessarily infinite e.g.

$$\lim_{n\to\infty} \frac{R}{n} = \Theta = P_r[z_1 \leq Z \leq z_2 | \mu,\sigma]$$

$$= \Phi\left(\frac{z_2-\mu}{\sigma}\right) - \Phi\left(\frac{z_1-\mu}{\sigma}\right). \tag{6.1}$$

Letting $z_2 \to \infty$ yields the special case discussed prevously.

Here the distribution function of Θ can no longer be as simply related to a single random variable β as in equations (3.7), (3.8) and the approximation given by (3.12). Further analysis of a more complex nature would be required to derive a reasonable approximate solution for the distribution of Θ.

The mean and variance of Θ, however, can easily be obtained. In fact

$$E(\Theta) = \Pr\left[-k_1\sqrt{\frac{N}{N+1}} \leq T \leq -k_2\sqrt{\frac{N}{N+1}}\right] = q \tag{6.2}$$

where T has the central student-t distribution with n-1 d.o.f.

Further

$$\mathrm{Var}(\Theta) = \Pr\left[z_1 \le Z_1 \le z_2,\ z_1 \le Z_2 \le z_2\right] - q^2 \tag{6.3}$$

where $Z_1-\bar{z}$ and $Z_2-\bar{z}$ have an exchangeable bivariate student distribution so that $T' = (Z_1-\bar{z},\ Z_2-\bar{z})$ has density

$$f(t) \propto \left(1 + \frac{Nt'S^{-1}t}{N^2-1}\right)^{-N/2}$$

where

$$S = s^2\begin{pmatrix} 1 & (N+1)^{-1} \\ (N+1)^{-1} & 1 \end{pmatrix}$$

For the exact probability function of R we need to calculate numerically the formidable integral (or a n-fold multivariate student integral).

$$\Pr[R=r] = \binom{n}{r}\int_0^{\infty}\int_{\infty}^{\infty}\left[\Phi\left(\frac{z_2-\mu}{\sigma}\right) - \Phi\left(\frac{z_1-\mu}{\sigma}\right)\right]^r\left[\Phi\left(\frac{\mu-z_2}{\sigma}\right) + \Phi\left(\frac{z_1-\mu}{\sigma}\right)\right]^{n-r} p(\mu,\sigma^2|\bar{z},s^2)\,d\mu\, d\sigma^2 \tag{6.4}$$

where $p(\mu,\sigma^2|\bar{z},s^2)$ is the posterior density of μ and σ^2. An approximation to the above based on the multivariate normal approximation to the multivariate student distribution proposed in section 3, is

$$\Pr[R=r] \doteq \binom{n}{r}\int_{\infty}^{\infty} h^r(y)[1-h(y)]^{n-r}\,d\Phi(y) \tag{6.5}$$

where

$$h(y) = \Phi\left(\frac{y}{\sqrt{N}} - k_2\sqrt{\frac{N-3}{N-1}}\right) - \Phi\left(\frac{y}{\sqrt{N}} - k_1\sqrt{\frac{N-3}{N-1}}\right) \tag{6.6}$$

This now involves only a single integral rather than the double integral.

The exact mean and variance of R/n can be obtained from the mean and variance of Θ since

$$E(R/n) = E(\Theta)$$

$$\mathrm{Var}(R/n) = \frac{n-1}{n}\mathrm{Var}(\Theta) + \frac{q(1-q)}{n}\ . \tag{6.7}$$

References

Aitchison, J. (1964), Bayesian tolerance regions, Journal of the Royal Statistical Society, B, 26, 161-175.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré 7, 1-68

Geisser, S. (1967). Estimation associated with linear discriminants. Annals of Mathematical Statistics, 38, 3, 807 817.

———— (1982). Aspects of the predictive and estimative approaches in the determination of probabilities, Biometric supplement, 38, 1, March 1982, 75-85.

Guttman, I. (1970), Statistical Tolerance Regions: Classical and Bayesian. Griffin, London.